

.tsp	XhoI	Start codon	
GTGTTAAATACGGCTCGAGGTTTAAAT		ATG TCT GTT GCC TTG TTA TGG GTT GTT TCT CCT TGT GAG	
		Met Ser Val Ala Leu Leu Trp Val Val Ser Pro Cys Asp	

Transit peptide of phytoene synthase			
GTG TCA AAT GGG ACA AGT TTC ATG GAA TCA GTC CGG GAG GGA AAC CGT			
Val Ser Asn Gly Thr Ser Phe Met Glu Ser Val Arg Glu Gly Asn Arg			

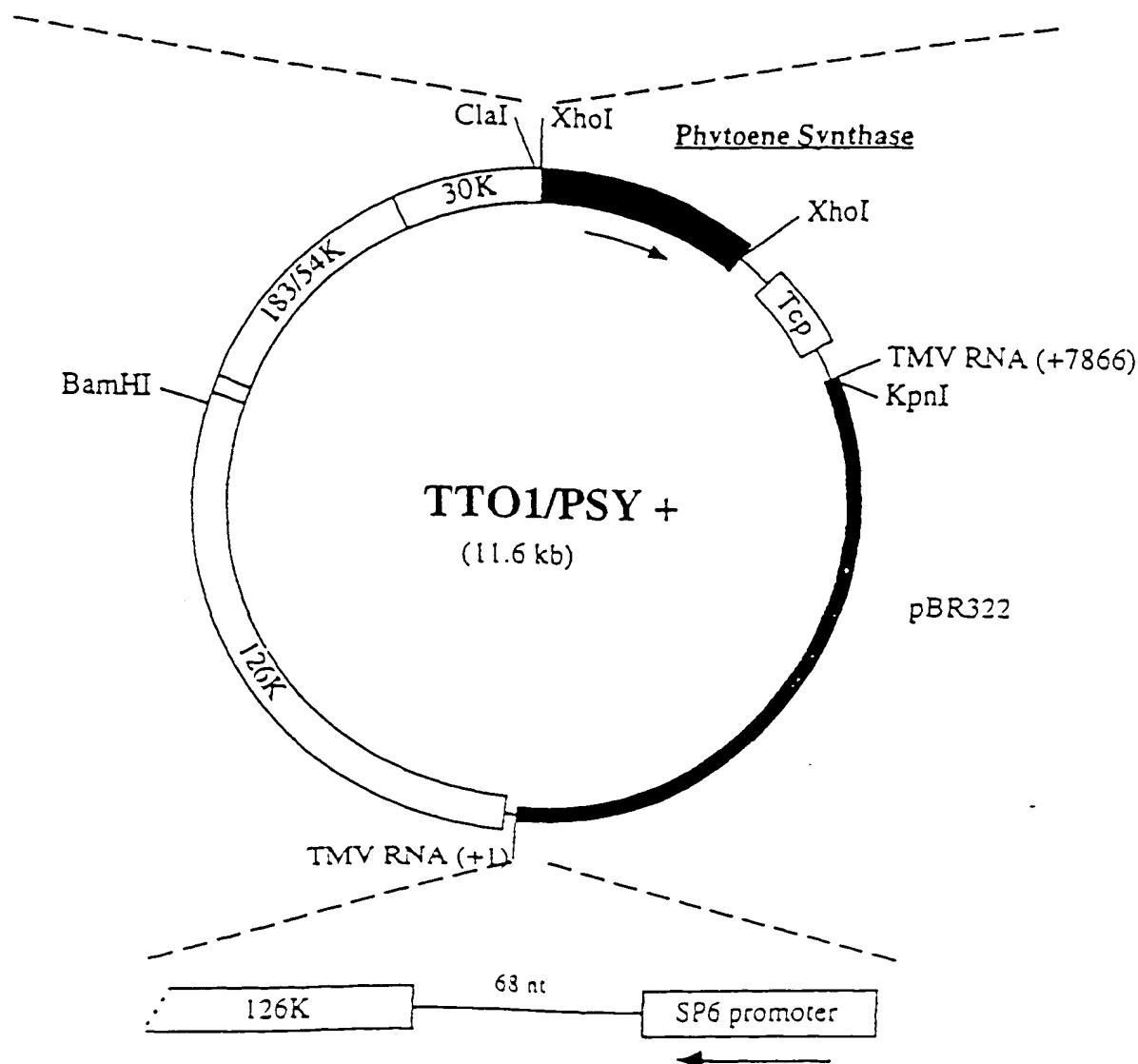


Figure 1

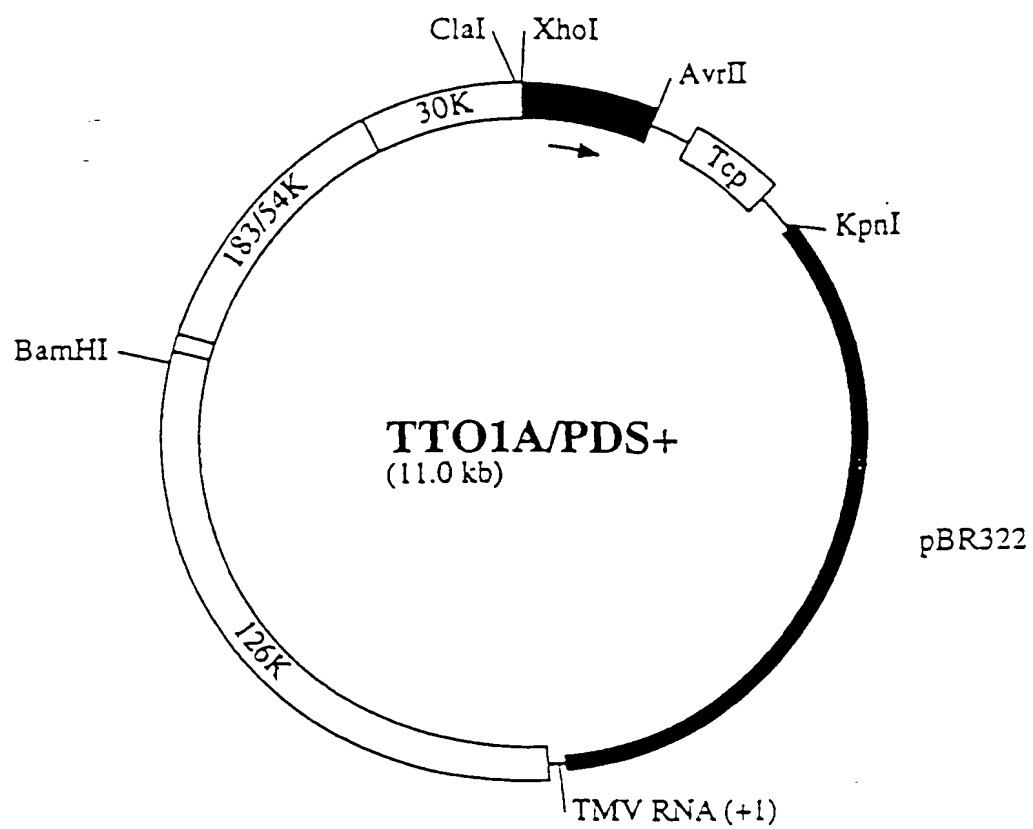


Figure 2

.tsp XhoI Start codon
 GTTTTAAATACGGCTGGAGTGCAGC ATG GAA ACC CTT CTA AAG CCT TTT CCA TCT CCT TTA CTT
 Met Glu Thr Leu Leu Lys Pro Phe Pro Ser Pro Leu Leu

Transit peptide of capsanthin-capsorubin synthase
 → TCC ATT CCT ACT CCT AAC ATG TAT AGT TTC AAA CAC AAC TTC ACT TTT
 Ser Ile Pro Thr Pro Asn Met Tyr Ser Phe Lys His Asn Ser Thr Phe

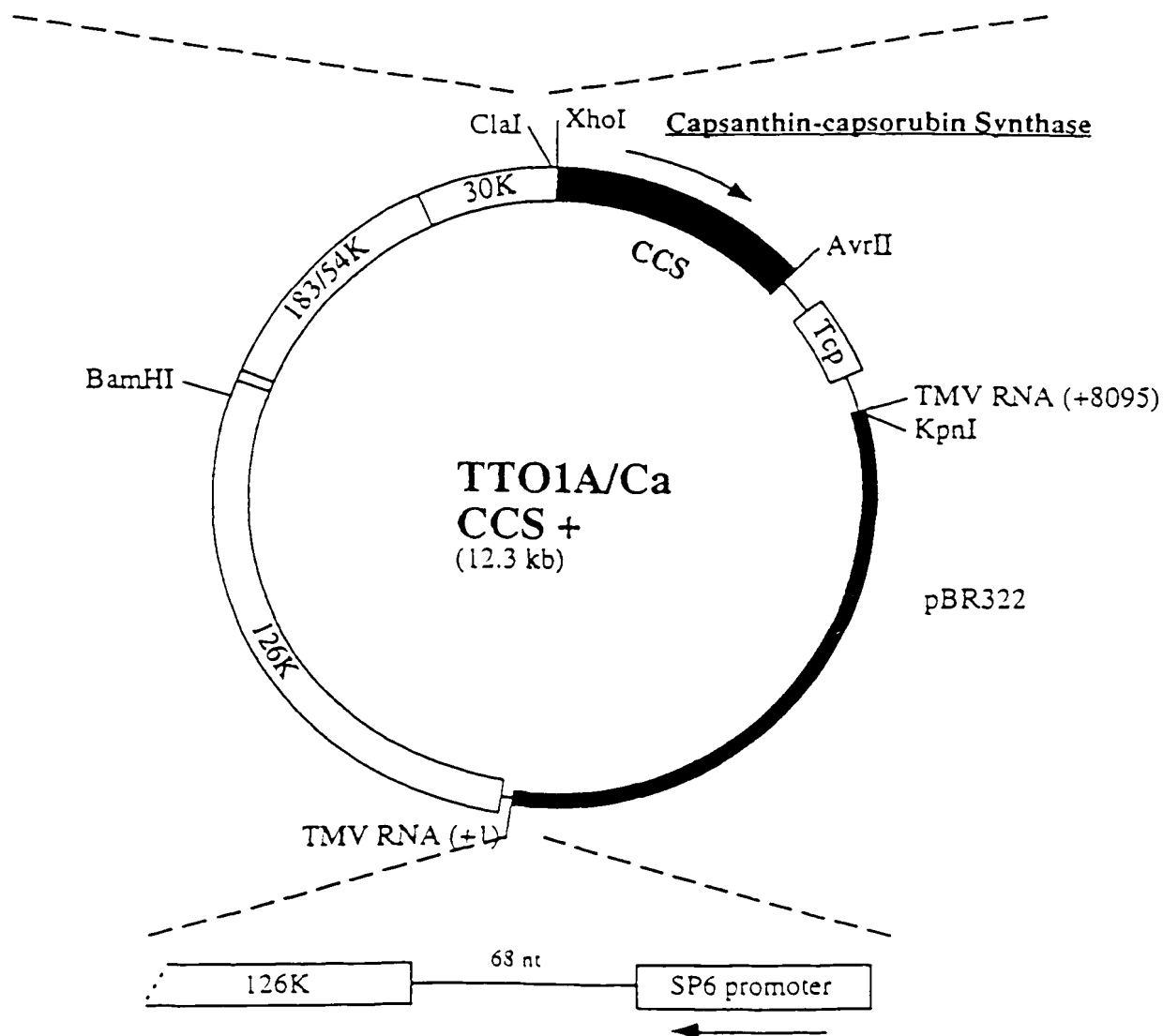


Figure 3

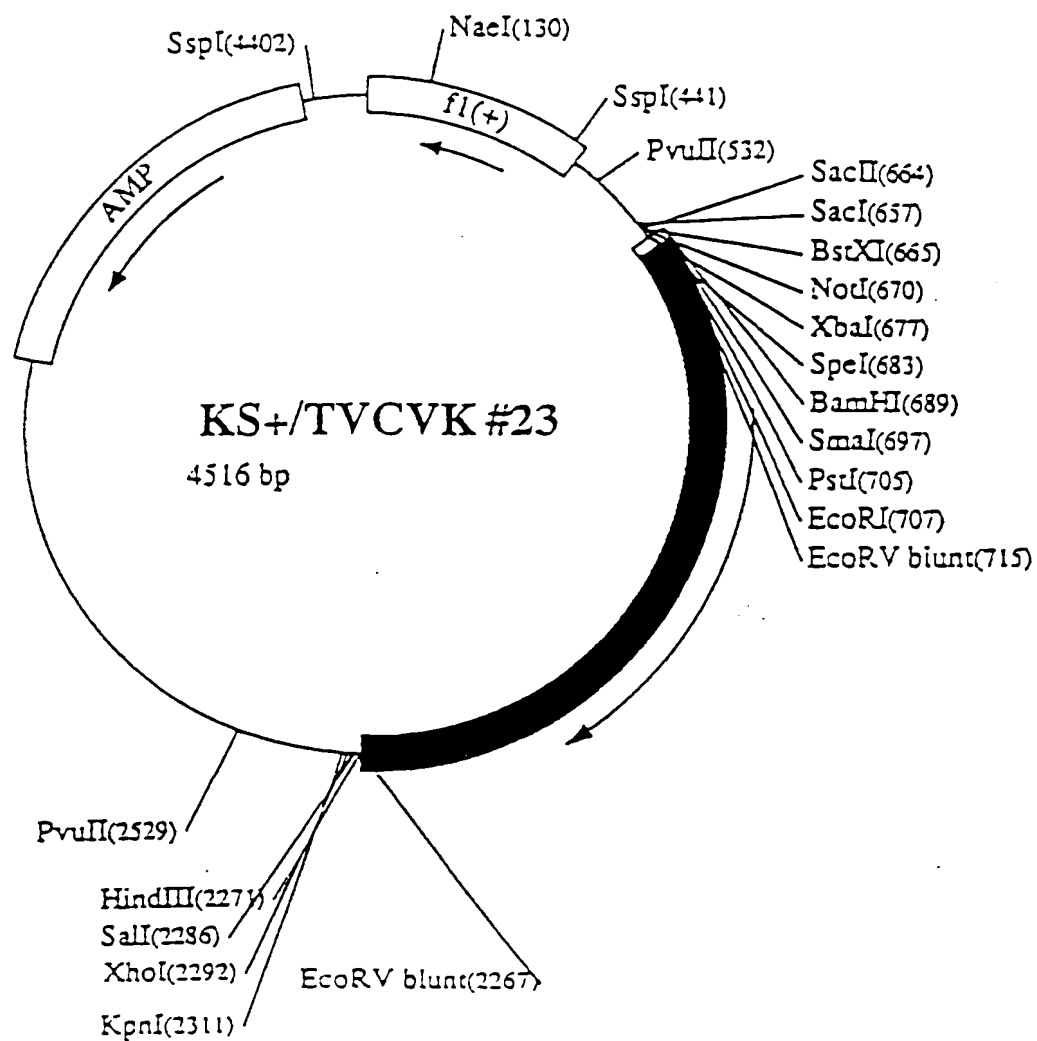


Figure 4

.bp XhoI Start codon
 GTTTTAAATACGCTCGAGCC ATG GCT TCC TCA GTT CTT TCC TCT GCA GCA GTT GCC ACC CGC
 Met Ala Ser Ser Val Leu Ser Ser Ala Ala Val Ala Thr Arg
 RUBISCO SSU Chloroplast Transit Peptide (*N. tabacum*)
 AGC AAT GTT GCT CAA GCT AAC ATG GTT GCA CCT TTC ACT GGC CTT
 Ser Asn Val Ala Gln Ala Asn Met Val Ala Pro Phe Thr Gly Leu

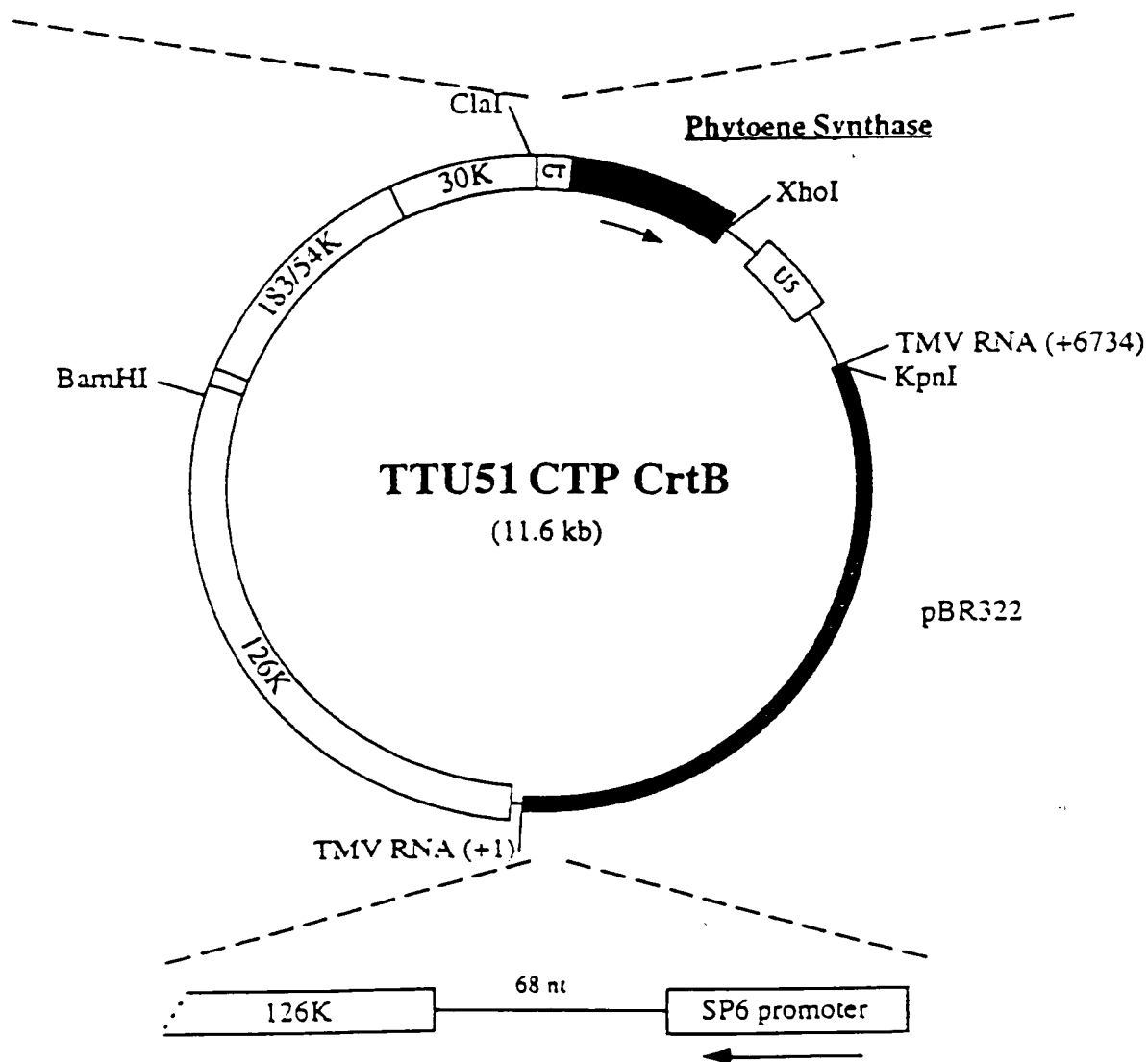


Figure 5

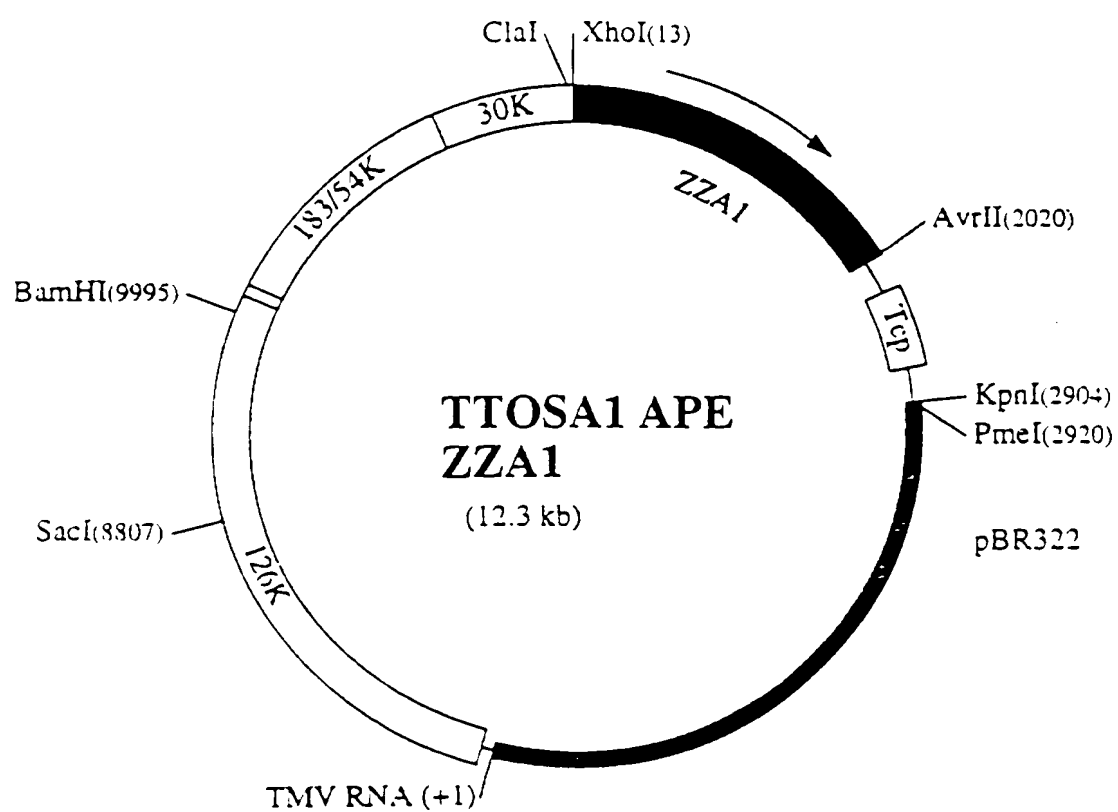


Figure 6

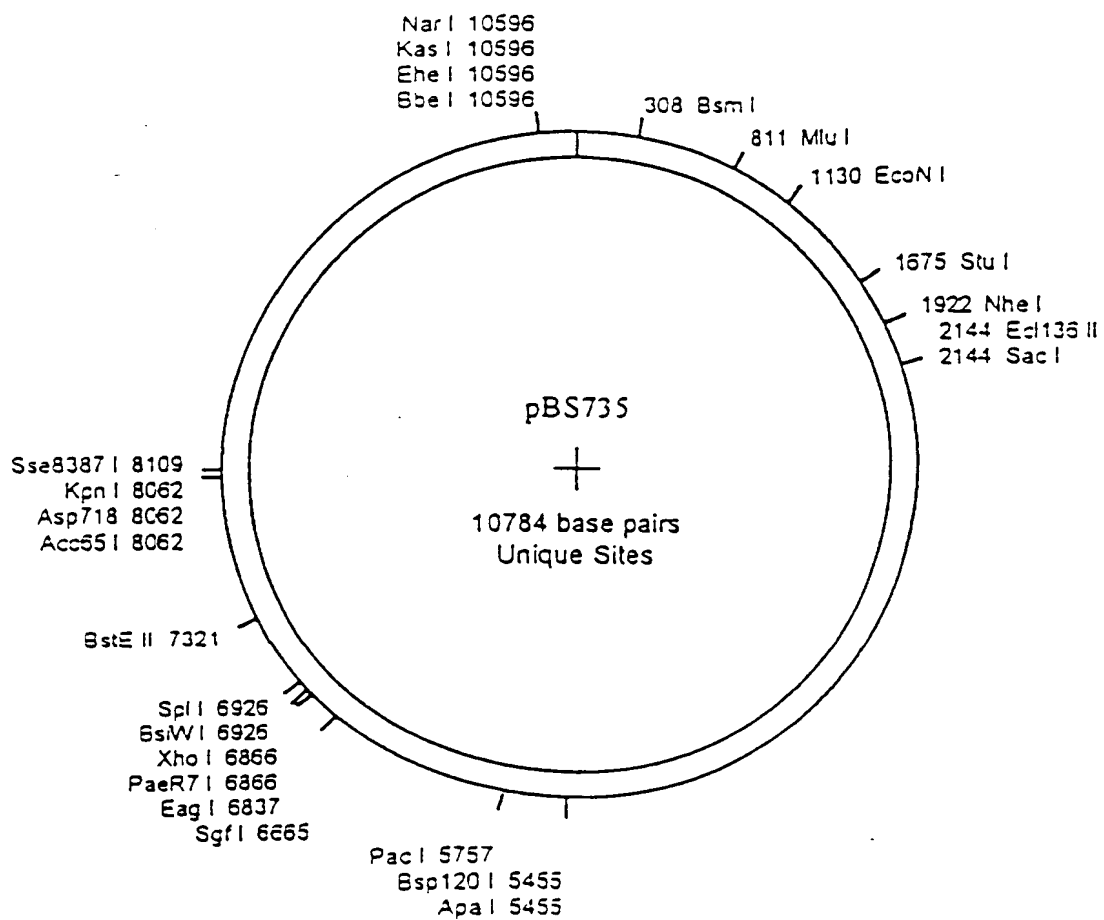


Figure 7

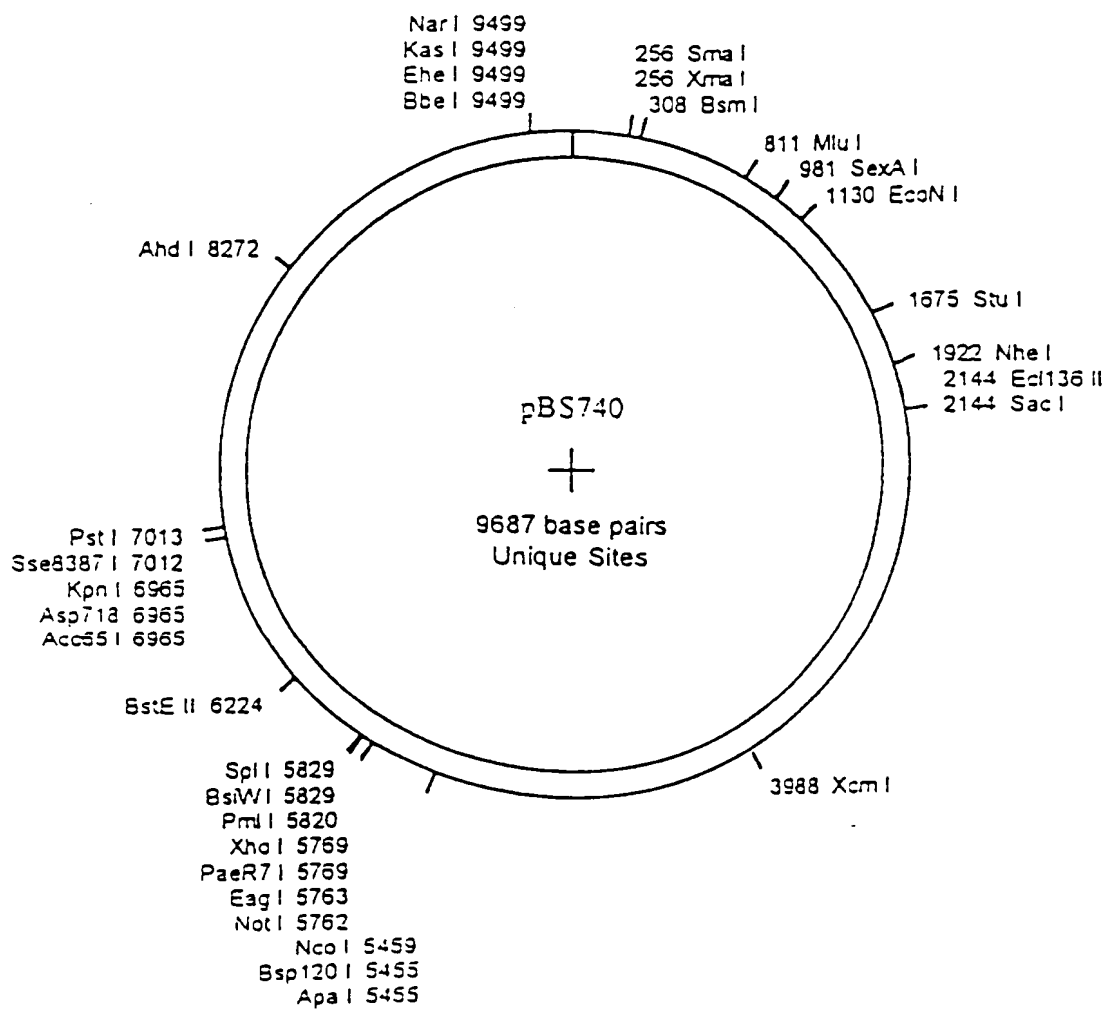


Figure 8

.usp XhoI Start codon
 GTTTAAATACGGTGGAGG ATG ATC AGA TTC TTA GTC CTC TCT TTG CTA ATT CTC ACC CTC TTA
 Met Ile Arg Phe Leu Val Leu Ser Leu Leu Ile Leu Thr Leu Phe

Signal peptide -1 +1 Mature α -trichosanthin
 CTA ACA ACT CCT GGT GTG GAG GGC | GAT GTT AGC TTC COT TTA TCA
 Leu Thr Thr Pro Ala Val Glu Gly | Asp Val Ser Phe Arg Leu Ser

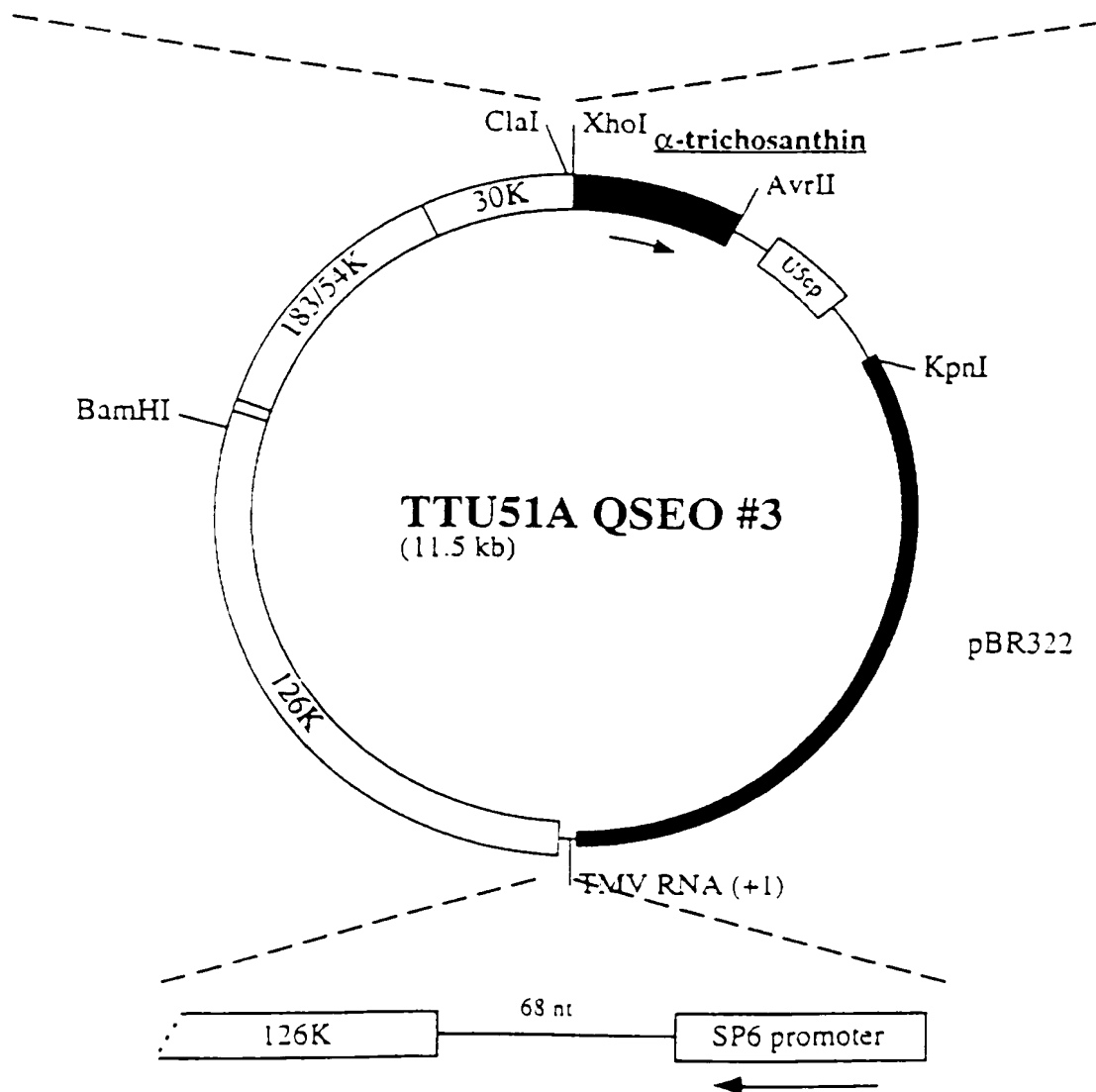


Figure 9

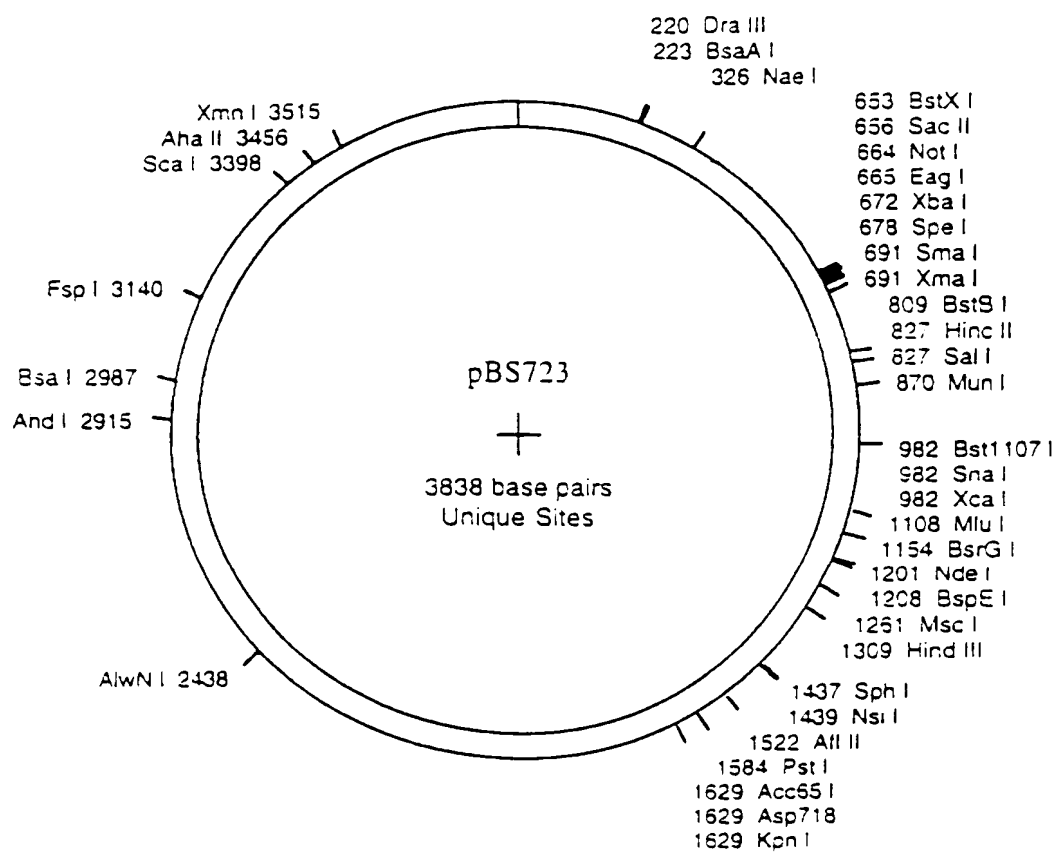


Figure 10

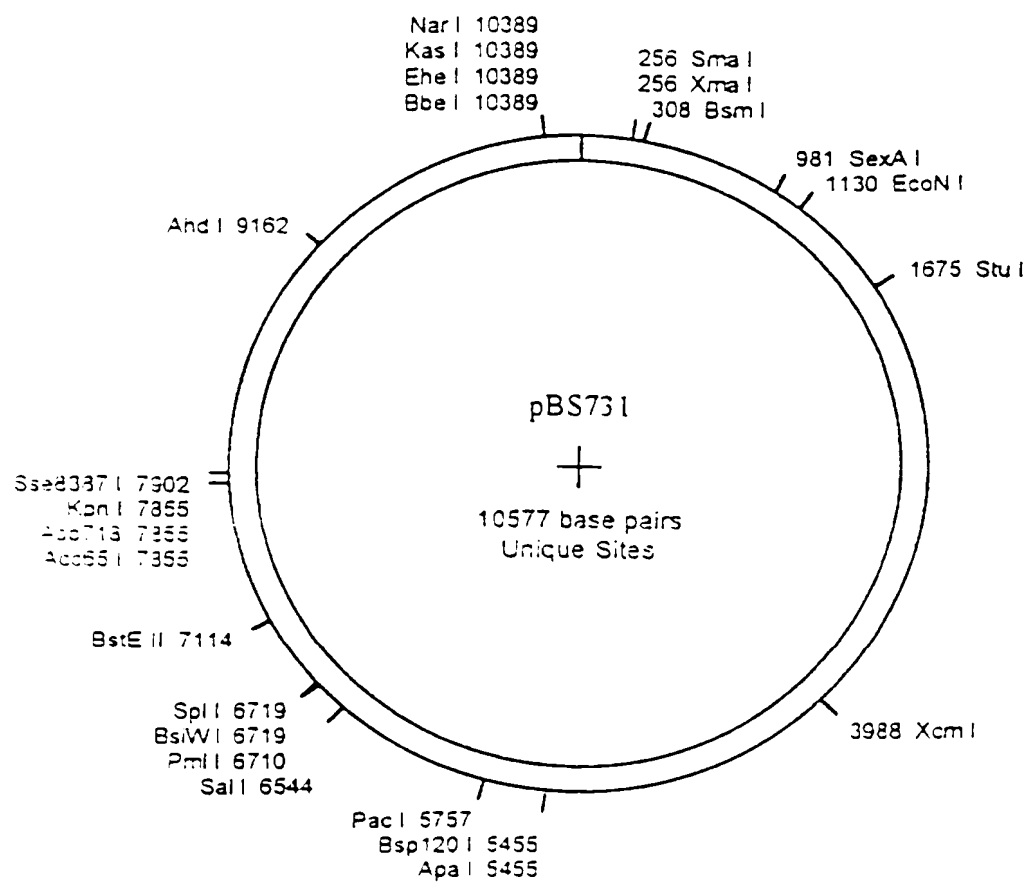


Figure 11

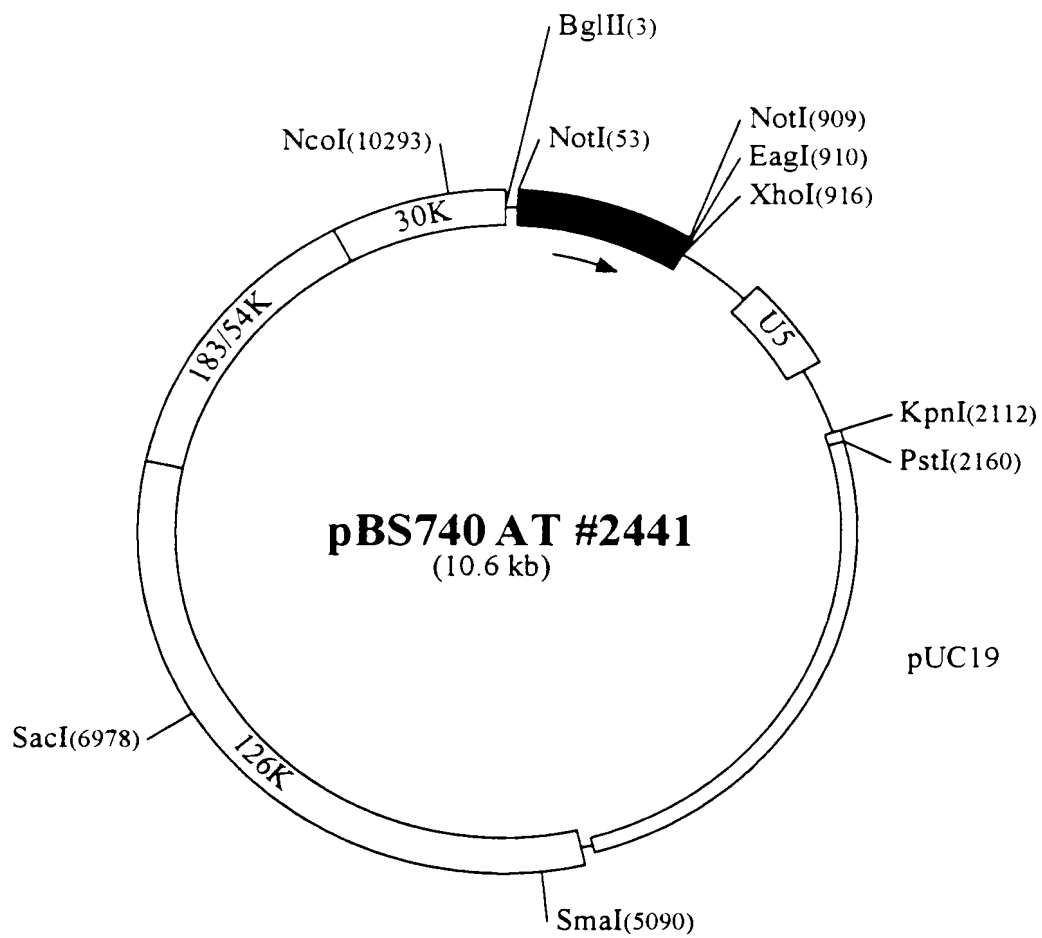


Figure 12

Nucleotide sequence of 740 AT #2441

ct tca ctt tct ccg

ATG GCT CTA CCT AAC CAG CAA ACC GTG GAT TAC CCT AGC TTC AAG CTC GTT ATC GTT GGC
M A L P N Q Q T V D Y P S F K L V I V G

GAT GGA GGC ACA GGG AAG ACC ACA TTT GTA AAG AGA CAT CTT ACT GGA GAG TTT GAG AAG
D G G T G K T T F V K R H L T G E F E K

AAG TAT GAA CCC ACT ATT GGT GTT GAG GTT CAT CCT CTT GAT TTC TTC ACT AAC TGT GGC
K Y E P T I G V E V H P L D F F T N C G

AAG ATC CGT TTC TAC TGT TGG GAT ACT GCT GGC CAA GAG AAA TTT GGT GGT CTT AGG GAT
K I R F Y C W D T A G Q E K F G G L R D

GGT TAC TAC ATC CAT GGA CAA TGT GCT ATC ATC ATG TTT GAT GTC ACA GCA CGA CTG ACA
G Y Y I H G Q C A I I M F D V T A R L T

TAC AAG AAT GTT CCA ACA TGG CAC CGT GAT CTT TGC AGG GTT TGT GAA AAC ATC CCA ATT
Y K N V P T W H R D L C R V C E N I P I

GTT CTT TGT GGG AAT AAA GTT GAT GTG AAG AAC AGG CAA GTC AAG GCC AAG CAG GTA ACA
V L C G N K V D V K N R Q V K A K Q V T

TTC CAC AGG AAG AAG AAC CTC CAG TAT TAC GAG ATA TCT GCC AAG AGC AAC TAC AAC TTC
F H R K K N L Q Y Y E I S A K S N Y N F

GAG AAG CCA TTC TTG TAC CTT GCT AGA AAG CTC GCC GGG GAC GCT AAT CTT CAC TTT GTG
E K P F L Y L A R K L A G D A N L H F V

GAA TCA CCT GCC CTT GCT CCC CCG GAA GTT CAA ATC GAC TTG GCT GCT CAG CAG CAG CAT
E S P A L A P P E V Q I D L A A Q Q Q H

GAG GCG GAG CTT GCA GCA GCA GCA AGT CAG CCA CTT CCT GAT GAC GAT GAT GAC ACC TTC
E A E L A A A A S Q P L P D D D D D T F

GAG TAG AGA AAG AGA GAT GTG ATC TGT CAC TGA TTA CCC GTT AGG GCT TGT CTG AAC TTT
E

TTT TTG TTC ATG GTG CTA TTT TTA TGT GTC CGT ACT TTG AAA TGA ATC GAT GAC ATT AGT

AAT TTT CAT TTT TAA GTT TTT AAC TGT CGC TAT GAA AGT GAA AAC

Figure 13

Nucleotide sequence alignment of 740 AT #2441 to AF017991
A. thaliana salt stress inducible small GTP binding protein Ran1

740 AT #2441	1	CTTCACTTTCGCCGATGGCTCTACCTAACAGCAAACCGTGGATTACCCCTAGCTTCAAGC	60
AF017991	67	CTTCACTTTCGCCGATGGCTCTACCTAACAGCAAACCGTGGATTACCCCTAGCTTCAAGC	126
740 AT #2441	61	TCGTTATCGTTGGCGATGGAGGCACAGGGAAGACCACATTTGTAAAGAGACATCTTACTG	120
AF017991	127	TCGTTATCGTTGGCGATGGAGGCACAGGGAAGACCACATTTGTAAAGAGACATCTTACTG	186
740 AT #2441	121	GAGAGTTTGAGAAGAAGTATGAACCCACTATTGGTGTGAGGTTTCATCCTCTTGATTTCCT	180
AF017991	187	GAGAGTTTGAGAAGAAGTATGAACCCACTATTGGTGTGAGGTTTCATCCTCTTGATTTCCT	246
740 AT #2441	181	TCACTAACTGTGGCAAGATCCGTTTCTACTGTTGGGATACTGCTGGCCAAGAGAAATTTG	240
AF017991	247	TCACTAACTGTGGCAAGATCCGTTTCTACTGTTGGGATACTGCTGGCCAAGAGAAATTTG	306
740 AT #2441	241	GTGGTCTTAGGGATGGTTACTACATCCATGGACAATGTGCTATCATCATGTTTGATGTCA	300
AF017991	307	GTGGTCTTAGGGATGGTTACTACATCCATGGACAATGTGCTATCATCATGTTTGATGTCA	366
740 AT #2441	301	CAGCACGACTGACATACAAGAATGTTCCAACATGGCACCGTGATCTTTGCAGGGTTTGTG	360
AF017991	367	CAGCACGACTGACATACAAGAATGTTCCAACATGGCACCGTGATCTTTGCAGGGTTTGTG	426
740 AT #2441	361	AAAACATCCCAATTGTTCTTTGTGGGAATAAAGTTGATGTGAAGAACAGGCAAGTCAAGG	420
AF017991	427	AAAACATCCCAATTGTTCTTTGTGGGAATAAAGTTGATGTGAAGAACAGGCAAGTCAAGG	486
740 AT #2441	421	CCAAGCAGGTAACATTCCACAGGAAGAAGAACTCCAGTATTACGAGATATCTGCCAAGA	480
AF017991	487	CCAAGCAGGTAACATTCCACAGGAAGGAGGAAGTCCAGTATTACGAGATATCTGCCAAGA	546
740 AT #2441	481	GCAACTACAAC TTCGAGAAGCCATTCTTGTTACCTTGCTAGAAAGCTCGCCGGGGACGCTA	540
AF017991	547	GCAACTACAAC TTCGAGAAGCCATTCTTGTTACCTTGCTAGAAAGCTCGCCGGGGACGCTA	606
740 AT #2441	541	ATCTTCACTTTGTGGAATCACCTGCCCTTGCTCCCCCGGAAGTTCAAATCGACTTGGCTG	600
AF017991	607	ATCTTCACTTTGTGGAATCACCTGCCCTTGCTCCCCCGGAAGTTCAAATCGACTTGGCTG	666
740 AT #2441	601	CTCAGCAGCAGCATGAGGCGGAGCTTGCAGCAGCAGCAAGTCAGCCACTTCCTGATGACG	660
AF017991	667	CTCAGCAGCAGCATGAGGCGGAGCTTGCAGCAGCAGCAAGTCAGCCACTTCCTGATGACG	726
740 AT #2441	661	ATGATGACACCTTCGAGTAGAGAAAGAGAGATGTGATCTGTCACTGATTACCCGTTAGGG	720
AF017991	727	ATGATGACACCTTCGAGTAGAGAAAGAGAGATGTGATCTGTCACTGATTACCCGTTAGGG	786
740 AT #2441	721	CTTGCTGAACTTTTTTTT	738
AF017991	787	CTTGCTGAACTTTTTTTT	804

Figure 14

Nucleotide sequence alignment of 740 AT #2441 to L16787
N. tabacum small ras-like GTP-binding protein

740 AT #2441	15	ATGGCTCTACCTAACCAGCAAACCGTGGATTACCCCTAGCTTCAAGCTCGTTATCGTTGGC	74
L16787	37	ATGGCTCTACCAAACCAACAACTGTAGATTATCCAAGCTTCAAGCTTGTAATCGTGGGC	96
740 AT #2441	75	GATGGAGGCACAGGGAAGACCACATTTGTAAAGAGACATCTTACTGGAGAGTTTGAGAAG	134
L16787	97	GATGGAGGAAC TGGGAAAACAAC TTTGTCAAGAGGCATCTTACTGGTGAATTTGAGAAG	156
740 AT #2441	135	AAGTATGAACCCACTATTGGTGTGAGGTTTCATCCTCTTGATTCTTCACTAACTGTGGC	194
L16787	157	AAATATGAACCCACTATTGGTGTGAGGTTTCATCCATTAGACTTCTTCACAAATGTGGG	216
740 AT #2441	195	AAGATCCGTTTCTACTGTTGGGATAC TCTGGCCAAGAGAAATTTGGTGGTCTTAGGGAT	254
L16787	217	AAAATTCGCTTTTATTGCTGGGATAC TCTGGACAAGAGAAGTTTGGAGGTCTTCGGGAT	276
740 AT #2441	255	GGTTACTACATCCATGGACAATGTGCTATCATCATGTTTGATGTACAGCAGGACTGACA	314
L16787	277	GGTTACTACATTCAATGGGCAATGCGCAATTATCATGTTTGATGTTACAGCCCGCTCGACC	336
740 AT #2441	315	TACAAGAATGTTCCAACATGGCACCGTGATCTTTGCAGGGTTTGTGAAAACATCCCAATT	374
L16787	337	TACAAGAATGTTCTTACGTGGCATCGAGATCTCTGCAGGGTTTGTGAAAACATCCCAATT	396
740 AT #2441	375	GTTCTTTGTGGGAATAAAGTTGATGTGAAGAACAGGCAAGTCAAGGCCAAGCAGGTAACA	434
L16787	397	GTTCTTTGTGGAACAAAGTTGATGTCAAGAACAGGCAAGTTAAGGCAAGCAAGTTACC	456
740 AT #2441	435	TTCCACAGGAAGAAGAACC TCCAGTATTACGAGATATCTGCCAAGAGCAACTACAAC TTC	494
L16787	457	TTCCACAGGAAGAAAAATTTGCAATACTATGAGATCTCAGCAAAGAGTAAC TACAAC TTT	516
740 AT #2441	495	GAGAAGCCATTCTTGTTACCTTGCTAGAAAGCTCGCCGGGACGCTAATCTTCAC TTTGTG	554
L16787	517	GAGAAGCCTTTTCTGTACCTTGCCAGAAAGCTTGCTGGGGATGCTAATCTTCAC TTTGTG	576
740 AT #2441	555	GAATCACCTGCCCTTGCTCCCCGGAAGTTCAAATCGACTTGGCTGCTCAGCAGCAGCAT	614
L16787	577	GAATCACCTGCAC TTTGCTCCCCCTGAAGTACAAATTGATTTAGCTGCACAGCAACTGCAT	636
740 AT #2441	615	GAGGCGGAGCTTG CAGCAGCAGCAAGTCAGCCACTTTCCTGATGACGATGATGACACCTTC	674
L16787	637	GAACAAGAGCTTTTGC AAGCCGCTGCGCACGC ACTTCCAGATGACGATGATGAAGCTTTT	696
740 AT #2441	675	GAGTAGA 681	
L16787	697	GAATAGA 703	

Figure 15

Amino acid sequence comparison of 740 AT #2441 to tobacco RAN-B1 GTP binding protein

Nt RAN-B1	MALPNQQTVDYPSFKLVIVGDGGTGKTTFVKRHLTGFEKKYEPTIGVEVHPLDFFTNCG
740 AT #2441	MALPNQQTVDYPSFKLVIVGDGGTGKTTFVKRHLTGFEKKYEPTIGVEVHPLDFFTNCG
Nt RAN-B1	KIRFYCWDTAGQEKFGGLRDGYIYGQCAIIMFDVTSTTDIQECSNMAP*SLQGL*KHSQ
	+ + + +
740 AT #2441	KIRFYCWDTAGQEKFGGLRDGYIYGQCAIIMFDVTARLTYKNVPTWHR-DLCRVENIP
Nt RAN-B1	LFFVGIKLM*KNRQVKAQ
	+ + +
740 AT #2441	IVLCGNKVDVKNRQVKAK

Figure 16

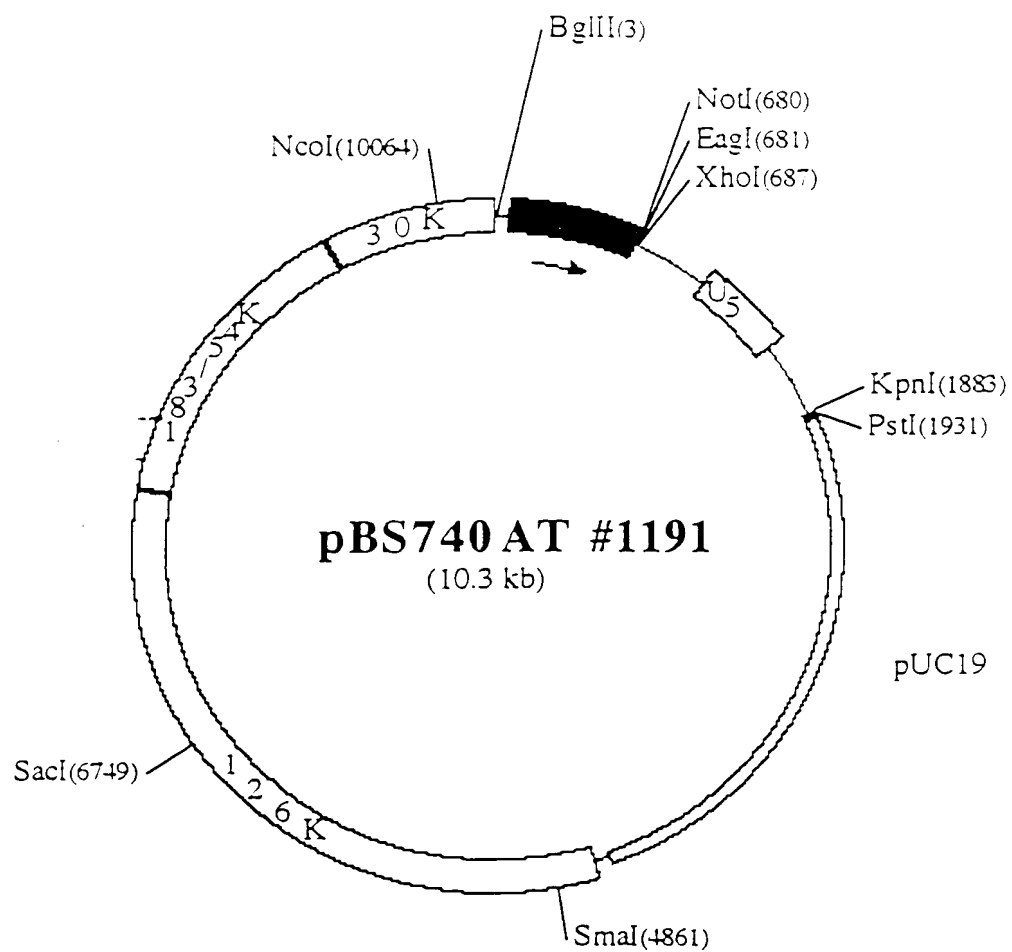


Figure 17

Nucleotide sequence of 740 AT #1191

GCT	ACT	ATG	GTT	GCC	TCT	CCG	GCT	CAG	GCC	ACT	ATG	GTC	GCT	CCT	TTC	AAC	GGA	CTT	AAG
A	T	M	V	A	S	P	A	Q	A	T	M	V	A	P	F	N	G	L	K
TCC	TCC	GCT	GCC	TTC	CCA	GCC	ACC	CGC	AAG	GCT	AAC	AAC	GAC	ATT	ACT	TCC	ATC	ACA	AGC
S	S	A	A	F	P	A	T	R	K	A	N	N	D	I	T	S	I	T	S
AAC	GGC	GGA	AGA	GTT	AAC	TGC	ATG	CAG	GTG	TGG	CCT	CCG	ATT	GGA	AAG	AAG	AAG	TTT	GAG
N	G	G	R	V	N	C	M	Q	V	W	P	P	I	G	K	K	K	F	E
ACT	CTC	TCT	TAC	CTT	CCT	GAC	CTT	ACC	GAT	TCC	GAA	TTG	GCT	AAG	GAA	GTT	GAC	TAC	CTT
T	L	S	Y	L	P	D	L	T	D	S	E	L	A	K	E	V	D	Y	L
ATC	CGC	AAC	AAG	TGG	ATT	CCT	TGT	GTT	GAA	TTC	GAA	GT							
I	R	N	K	W	I	P	C	V	E	F	E	V							

Figure 18

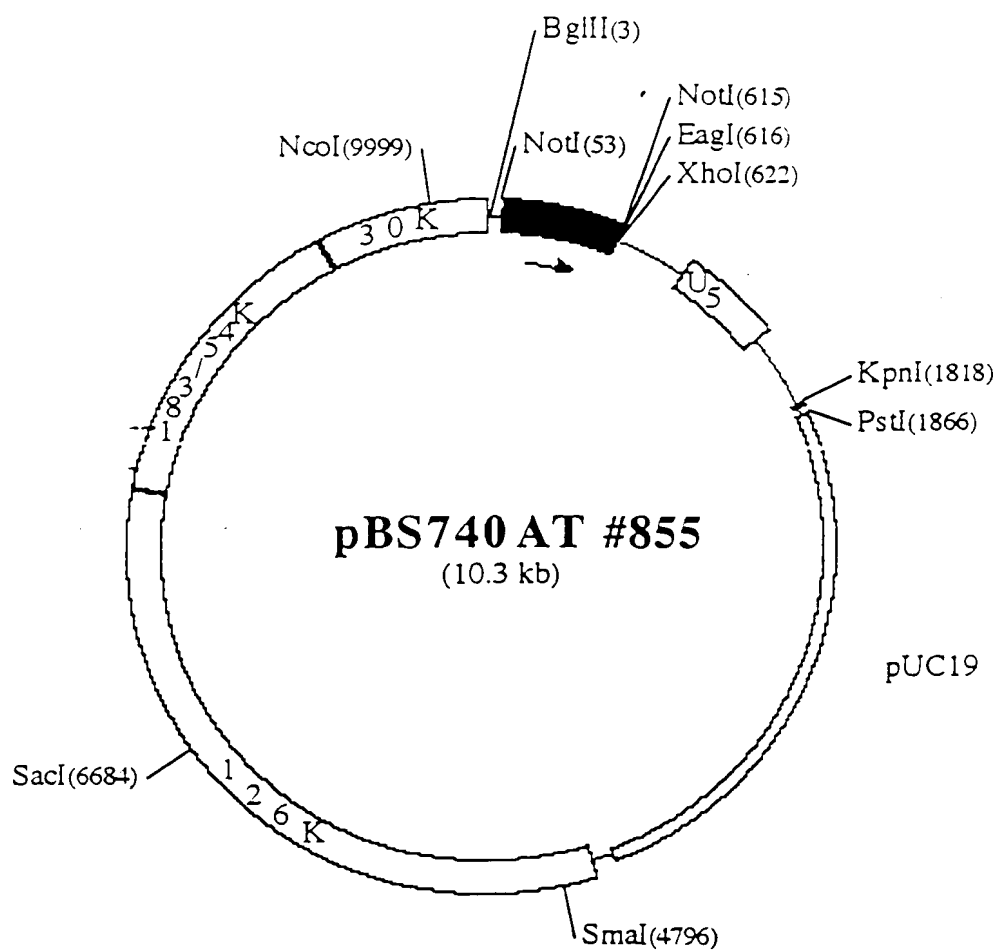


Figure 19

Nucleotide sequence alignment of 740 AT #855 to *Arabidopsis thaliana*
HAT7 homeobox protein ORF (U09340)

HAT7	GAA ACG ACG TCG GCT AGT TAT TGG GCA TGG CCT GAC CAG CAG CAA
	E T T S A S Y W A W P D Q Q Q
HAT7	CAA CAT CAC AAT CAT CAT CAG TTC AAT TGA tcatattgtctaagaaca
	Q H H N H H Q F N *
HAT7	acatcatactcatcttgatatcattattatcatcaaaagaaaattccgtagattttttta
HAT7	ataagtattttcaaattatttggcacgttttaaaattaattaaattgggttattatgttt
740 AT #855	acttgatctgtttcataactaaaaccaaaactcatgtttgttcactccaaacacaaacaca
	--
HAT7	- acttgatctgtttcataactaaaaccaaaactcatgtttgttcactccaaacacaaacaca
740 AT #855	gcagtaatcaaaaatcgtcttataacaaaaaggaaatgcaacaaaacagaagaaacaact
HAT7	gcagtaatcaaaaatcgtcttataacaaaaaggaaatgcaacaaaacagaagaaacaact
740 AT #855	aagtagtaggcaagattcttcttcactcgtcttcttggctacggagcc
HAT7	aagtagtaggcaagattcttcttcactcgtcttcttggctacggagcc

Figure 20

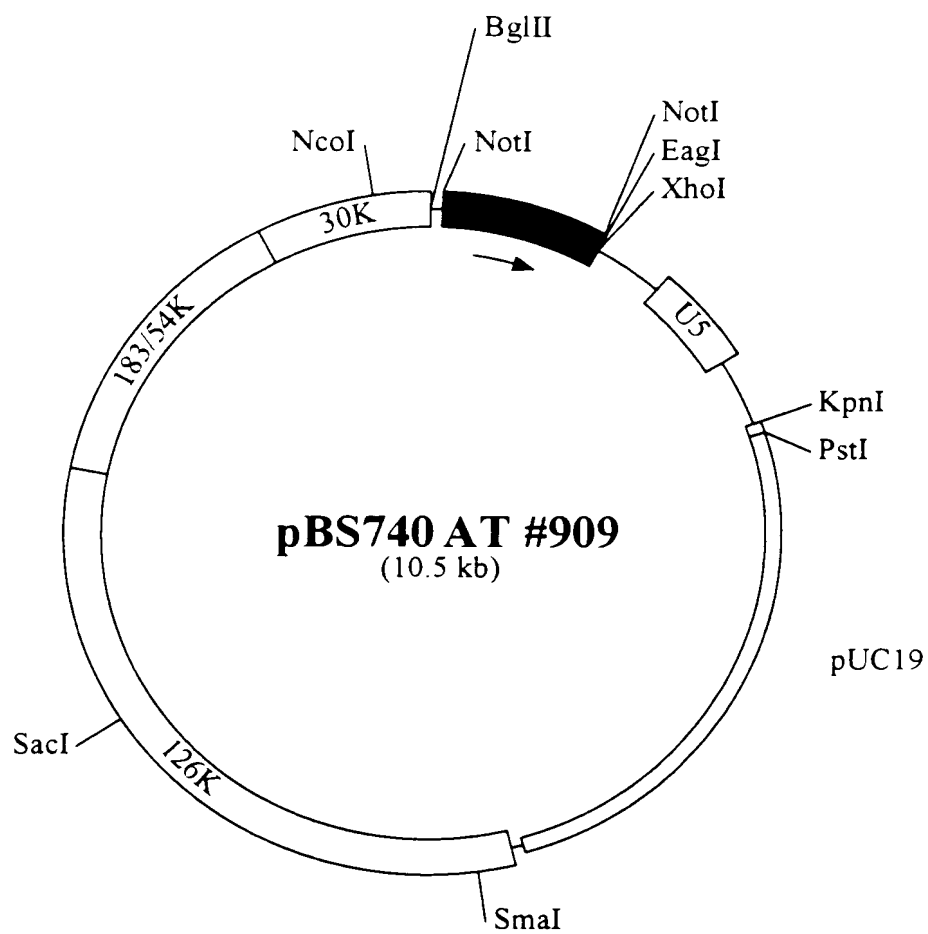


Figure 21

Nucleotide sequence alignment of 740 AT #909 to *H. sapiens* S56985 ribosomal protein L19
[human, breast cancer cell line, MCF-7]

740 AT #909	1	GAAGCGGCTCGCCGCATCAGTGATGAAGTGCGGGAAGGGCAAAGTTTGGCTCGATCCCAA	60
S56985	32	GAAGAGGCTCGCCTCTAGTGTCTCCGCTGTGGCAAGAAGAAGGTCTGGTTAGACCCCAA	91
740 AT #909	61	CGAAAGCTCCGACATCTCCATGGCCAATTCCCGCCAAAACATCAGGAAGCTTGTGAAGGA	120
S56985	92	TGAGACCAATGAAATCGCCAATGCCAACTCCCGTCAGCAGATCCGGAAGCTCATCAAAGA	151
740 AT #909	121	TGGTTTCATCATCAGGAAGCCAACCAAGATTCACTCTCGTTCAGAGCTCGCAAATGAA	180
S56985	152	TGGGCTGATCATCCGCAAGCCTGTGACGGTCCATTCCCGGGCTCGATGCCGGAAAAACAC	211
740 AT #909	181	GATTGCCAAGATGAAGGGTCGTCACCTCTGGATACGGTAAGAGGAAGGGTACCCGTGAAGC	240
S56985	212	CTTGCCCCGCCGAAGGGCAGGCACATGGGCATAGGTAAGCGGAAGGGTACAGCCAATGC	271
740 AT #909	241	TAGGTTGCCAACAAAGGTACTGTGGATGCGTAGGATGCGTGTTCTTAGGCGTCTGTTGAA	300
S56985	272	CCGAATGCCAGAGAAGGTCACATGGATGAGGAGAATGAGGATTTTGCGCCGGCTGCTCAG	331
740 AT #909	301	GAAATACAGAGAGACGAAGAAGATTGACAAGCACATGTACCATGACATGTACATGCGTGT	360
S56985	332	AAGATACCGTGAATCTAAGAAGATCGATCGCCACATGTATCACAGCCTGTACCTGAAGGT	391
740 AT #909	361	TAAGGGTAATGTGTTCAAGAACAAGCGTGTCTTGATGGAGAGTATCCACAAGTCAAAGGC	420
S56985	392	GAAGGGGAATGTGTTCAAAAACAAGCGGATTCTCATGGAACACATCCACAAGCTGAAGGC	451
740 AT #909	421	TTAGAAGCTAGGGGAGAA	438
S56985	452	AGACAAGGCCCGCAAGAA	469

Figure 22

Amino acid sequence alignment of 740 AT #909 to human P14118 60S ribosomal protein L19

740 AT #909	1	KRLAASVMKCGKGVWLDPNESDISMANSRQNIRKLVKDGFIIRKPTKIHSR SRARKMK	60
		KRLA+SV++CGK KVLDPNE+++I+ ANSRQ IRKL+KDG IIRKP +HSR+R RK	
Human P14118	8	KRLASSVLRCGKKKVWLDPNETNEIANANSRQQIRKLIKDGLIIRKPVTVHSRARC RKNT	67
740 AT #909	61	IAKMKGRHSGYGKRKGTREARLPTKVLWMRRMRVLRLLKKYRETKKIDKHMYHDMYMRV	120
		+A+ KGRH G GKRKGT AR+P KV WMRRMR+LRLL++YRE+KKID+HMYH +Y++V	
Human P14118	68	LARRKGRHMGIGKRKGTANARMPEKVTWMRRMRILRLLRRYRESKKIDRHMYSLYLKV	127
740 AT #909	121	KGNVFKNKRVLMESIHKSKA*KLGEK	146
		KGNVFKNKR+LME IHK KA K +K	
Human P14118	128	KGNVFKNKRILMEHIHKLKADKARKK	153

Figure 23